图书情报工作
第63卷 第15期 2019年8月

学科领域科研产出的空间分布规律研究*

——以计算机软件与应用学科为例

■ 马超 李纲 毛进 谷岩松
武汉大学信息资源研究中心 武汉 430072

摘要：[目的/意义]探究学科领域科研产出在空间上的分布情况，发掘其空间分布是否存在自相关性，分析其形成空间聚集的影响因素。[方法/过程]基于计算机领域核心期刊从1997－2016年间的题录数据，提取机构地理位置信息，以此为观测点，对其分布情况进行可视化呈现，并测算基尼指数和中心性指数，研究其分布集中情况；通过测算全局莫兰指数和局部自相关指标，研究其空间自相关规律；通过测算皮尔森相关系数，研究其与地区高校数量、国民生产总值、研究与试验投入和行业从业人员数量的相关性。[结果/结论]计算机领域核心期刊论文产出呈现出地理位置分布不均衡，有明显集中态势，且存在空间自相关性，同时与地区该学科高水平高校数量、行业从业人员数量和科研人员数量相关性较高。

关键词：空间知识溢出 空间自相关 城市创新力 探索性空间数据分析
分类号：G322
DOI：10.13266/j.issn.0252－3116.2019.15.011

随着研究方法的革新与跨学科研究的不断发展，研究机构和高校的科研水平不断提高，研究成果的数量与质量都在不断攀升。但是由于区域发展的不均衡性，不同地区之间的差异较为明显。这种不均衡性引起了许多关于科研产出空间分布规律的研究。如L. Bornmann等学者通过构建密度地图探测SCI中的科研产出“热区”[1]；C. W. Matthiessen等学者曾基于SCI中的文章数量来分析欧洲地区的科研产出情况[2]；P. Zhou等学者基于SCI中的文章数量来分析中国各省份的科研产出情况[3]。结合科研产出在地理地图上的可视化呈现，其分布情况和集中情况得到了充分的展示与研究。

与此同时，一些研究尝试探索科研产出分布不均的影响因素。其中，学术系统的内部因素被认为是主要影响因素，如M. Mabe论证了研究人员数量和科研产出有非常直观的正相关关系[4]。地理位置等往往被作为客观条件，但其本身作为影响因素被考虑的较少。合作作为一种提高科研产出效率与质量的形式，不仅仅局限于机构之内，随着科学技术水平的不断发展，跨机构、跨地区的合作行为也被广泛采纳。有学者也基于合作网络在地图上进行了可视化呈现[5]。然而，在跨地区合作的过程中，势必会受地理位置等影响。这种影响通常表现为地理位置接近的科研合作机构开展的科研合作更为频繁，呈现出空间上聚集性。那么这种源自于地理位置因素的影响，是否仅仅影响了合作网络，或者是间接地影响了其他的因素，从而对科研产出的分布情况产生影响？根据地理学第一定律[6]，任何事物都存在相关性，且地理位置相近的事物联系更加紧密。这种聚集性和关联性即为空间上的自相关性[7]。由于空间自相关的普遍存在，从地理学的角度来对科研产出进行空间分布规律的分析，具有很高的研究价值。此外，科研产出在空间上的聚集也可能会受地区发展水平的影响，并且这样的影响也具有相互性，即科研的产出的集中分布会对区域的经济发展、创新能力产生影响，这种影响常被概括为空间知识溢出(knowledge spillover)[8]。这样的影响主要来自于哪些因素，其影响程度如何，也是值得分析与挖掘的研究问题。

* 本文系国家自然科学基金重大课题“国家安全大数据综合信息集成与分析方法”(项目编号：71790612)和中国博士后科学基金项目“融合语义与关系的科研社群识别与演化研究”(项目编号：2018M630885)研究成果之一。
作者简介：马超(ORCID：0000－0002－3425－993X)，博士研究生；李纲(ORCID：0000－0001－5573－6400)，主任，教授，博士生导师；毛进(ORCID：0000－0001－9572－6709)，博士后，通讯作者，E-mail：danveno@163.com；谷岩松(ORCID：0000－0001－9403－7442)，博士研究生。
收稿日期：2018－12－28 修回日期：2019－04－02 本文起止页码：94－102 本文责任编辑：徐健

笔者将针对计算机软件与应用领域的科研分布情况进行空间层面的揭示与探究，不仅能够直观地得到该领域科研产出的空间分布情况，并能够从外因角度分析该领域科研分布不均产生的原因，对于发现区域科研发展共性，研究区域科研发展方向具有显著意义。笔者将基于计算机软件与应用领域核心期刊从1997年到2016年20年间论文产出的题录数据，采用统计分析与探索性空间数据分析(ESDA)研究方法，结合地理空间可视化，探究并揭示其中的时空分布与空间自相关性，并对其影响因素进行探讨。

1 相关研究

1.1 空间科学计量

最早将空间分析方法用于科研产出方面的研究论文发表于20世纪70年代，但其后并未得到长足发展，近年来该领域才逐渐得到关注。2009年，K. Frenken等学者将与科学地理空间相关的科学计量学研究定义为空间科学计量学(Spatial Scientometrics)[9]，同时将空间科学计量学的研究内容分为3类：空间分布情况(spatial distribution)、空间偏见(spatial biases)以及引文影响力(citation impact)。其中，对于科学空间分布情况的相关研究主要基于国家层面，而更加深入的基于地区层面的研究较少，其主要瓶颈在于地理位置信息的获取难度与精度，其研究方法主要为进行可视化呈现，并进行描述性分析。空间偏见相关研究主要集中于科研产出机构在空间上的聚集性，研究主要集中在其背后的原因探究。这种集中性至少可以从3种角度来解释：①空间上互相接近的科研主体存在交集的概率更高；②空间距离和面对面沟通交流的成本成正比，如交通费用、通行时间等；③空间上接近的主体在“科研风格”上更为接近。空间分布情况和空间偏见是空间科学计量学中主要的研究主题。

1.2 城市科研产出影响因素

关于城市科研产出影响因素的相关研究较少，但由于科研产出存在空间溢出现象，使得城市科研产出对城市经济发展产生影响，因此通常依托于关于城市创新能力的研究。有关于知识空间溢出现象，论文和专利作为科研产出的重要形式，在以城市为单位的研究中，被视为新型城市创新力和竞争力的根源[10]，也经常被作为评价城市创新力的重要指标[11－12]。而形成城市创新力差异的影响因素有很多，如曹勇等以城市经济发展水平、对外交流能力、科研投入水平、成果转化能力为一级指标，重点分析了城市人口总数、城市国民生产总值(GDP)以及城市研究与试验(R&D)资源投入占比对我国4个直辖市科研产出能力差异原因进行了比较分析[13]，其中最为关键的是研究与试验(R&D)资源投入。有很多学者从不同角度对科研投入与科研产出之间的关系进行了分析与研究。史欣向等基于R&D资源清查数据分析了研发资源与科研产出的关系[14]，杨军、吴杨等也分别探讨了高校科研投入与产出之间的关系[15－16]。同时，高校作为科研工作的主要阵地，依托高校而建设的研究所、实验室、研究中心对高校所在城市的科研产出也起到至关重要的影响，因此高校数量也应作为城市综合科研产出影响因素。

前者的研究主要集中在传统计量学方法的应用和科研合作网络的呈现与分析，主要研究意义在于科研合作网络内部规律的揭示，探究内部成因。而从空间计量角度进行的空间统计分析，基于知识溢出现象的外部影响因素探究的相关研究相对缺乏。结合前者研究，本文的主要研究内容有2个方面：①可视化展示并分析特定学科领域的科研产出的空间分布情况，并揭示其分布规律；②结合空间分布情况，比较并分析科研产出集中地区的形成原因以及相关因素分析。

2 研究方法

2.1 数据来源

本文实验数据来源于中国知网。首先，利用期刊检索功能，选择计算机软件及计算机应用分类下的核心期刊，共包括《软件学报》等10种期刊；然后，针对这些期刊导出1997年至2017年间包含机构的完整题录信息，共得到78 112条记录。去除1 954条无效数据，如征稿启事、会议通知等，最终得到有效数据76 113条，其中包含题目、作者、机构、关键词、摘要等信息。

2.2 机构与地址解析

数据清洗阶段的关键问题是入库数据格式标准不一致问题。如1997年的导出文件中的机构信息仅包含机构全称，并且以逗号“,”为分隔符，而2002年的导出文件中机构信息以“机构全称，机构所在城市，邮政编码”的格式存储，同时又以感叹号“!”为分隔符，而2007年以后的导出文件中机构又以分号“;”为分隔符。同时不同期刊不同年份的机构字段也略有差异。这对数据的批量处理产生了极大的阻碍。本阶段采用字符串替换与截取，结合正则匹配方法进行数据清洗，从原始数据中提取出完整的机构全称。

数据处理主要包含两个阶段的任务，第一阶段为

机构顶级名称提取，第二阶段为地址解析。具体的数据处理流程如图 1 所示：

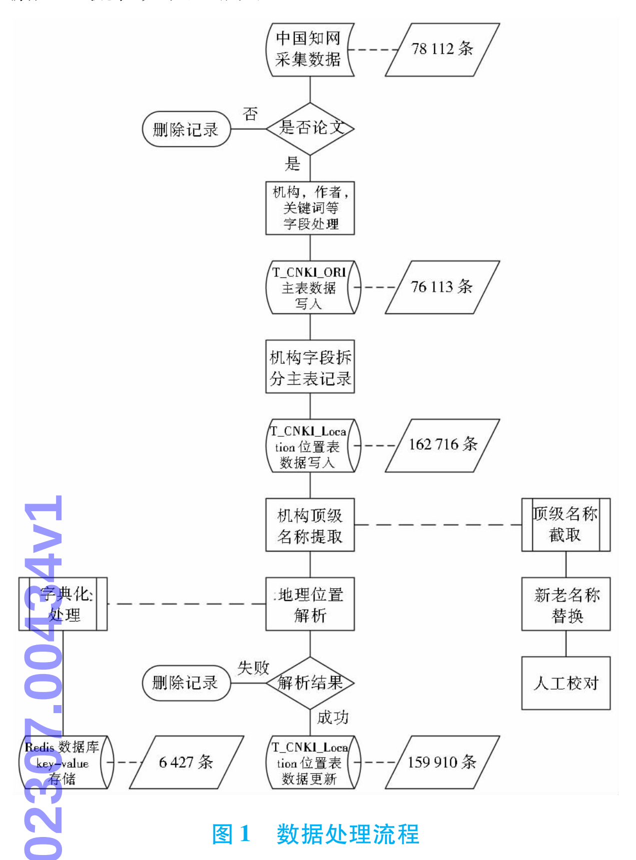


图 1　数据处理流程

第一阶段采用关键词匹配结合字符串截取方法，提取机构顶级名称。共获得 162 716 条机构顶级名称数据。

第二阶段为数据处理关键阶段，其主要目的是获取机构的地理位置坐标以及其所属的区县级、地级、省级行政区，用于后期的数据分析。首先，根据具体机构名称，使用高德地图 API 提供的地址解析（地理位置编码）功能，按照文档要求封装请求并提交；然后，进行返回的 JSON 数据解析，从完整的数据中提取省份、城市、经度和纬度，并存入数据库。同时将成功的解析结果存入内存数据库 Redis，供后期解析使用。由于部分机构名称无法进行地址解析，最终得到有效数据 159 910 条。

数据处理阶段有 2 个关键问题：①机构层次不一致。机构全称字段主要包含大学及学院、研究所、公司、集团等，如机构类型为大学时，有些条目精确到二级学院，而有些则只精确到大学名称。首先以“大学”为关键词进行匹配，截取科研机构为大学的机构顶级名称，如武汉大学等；然后只针对剩余数据进行“学院”关键字匹配，排除机构全称为“XX 大学 XX 学院”的歧义问题。最后进行“学校”“公司”等剩余关键字的匹配。②地址解析能力不足。引发该问题的主要原因有 3 个：一是时间跨度大，机构名称准确度不够，部分机构存在名称变更、机构合并，或者机构取消等关键变化；二是机构名称粒度过细，如机构名称为“XX 实验室”“XX 中心”时，地址解析效果不够理想；三是机构名称的多词一义，如坐落于安徽合肥的“中国科学技术大学”[17]，前者为其官方名称，但实际处理过程中存在“中科大”“中国科大”“中国科学与技术大学”“中国科技大学”等多个非官方名称，这些名称均无法进行地址解析。

针对该问题，笔者采用以下方法解决。首先进行人工校准，将无法匹配的旧名称通过百度百科和搜索引擎，查询机构新名称以及所在城市，并替换为新名称。同时在进行地址解析时，将机构名称与机构所在城市同时作为参数提交，提高解析能力。对于多词一义问题，在人工校准的同时，采用构建数据字典的方式，进行增量匹配，提高作业效率。

2.3　科研产出空间分析方法

笔者主要从时空维度探讨科研产出的演变，主要研究内容围绕科研产出机构在同一时间内空间上的分布和变化情况以及所在段时间内的变化情况。使用 Python 进行数据处理与基础分析，使用开源时空数据分析工具 GeoDa 进行探索性空间分析。

2.3.1　集中程度分析　D. B. Audretsch 等利用基于地理位置因素的基尼指数（Gini Index）来度量空间维度上的集中程度，同时也计算了中心性指数（CR_3）[18]。基于位置信息的基尼指数的计算公式如下：

$$G = \frac{1}{2n^2 \bar{Z}} \sum_{i=1}^{n} \sum_{j=1}^{n} |z_i - z_j| \qquad \text{公式(1)}$$

其中，n 代表科研产出机构所在地区的总数，$\bar{Z}$ 代表科研产出机构所在地区总平均数，z_i 代表地区 i 的科研产出数，j 和 i 类似。根据不同的观测点选取，可以分别以省级、地级、县级或是直接以机构进行地区划分。

中心性指数的计算公式如下：

$$CR_3 = \sum_{i=1}^{3} w_i \qquad \text{公式(2)}$$

其中，w_i 代表在同一时期内科研产出最多的机构所在地区的前 3 名的科研产出数量除以总产出。

2.3.2　空间相关性分析　在探索性空间分析阶段，主要进行空间基础统计分析和空间自相关分析。空间自相关是指给定变量在同一个分布区内的观测数据之间潜在的相互依赖性，包含全局空间自相关和局部空间自相关。全局空间自相关用于识别整体的空间自相关

情况，而局部空间自相关用于发现特殊的空间关系[19]。

莫兰指数(Moran's Index)常用于描述全局空间自相关。其计算公式如下：

$$I_t = \frac{N}{S}\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}W(i,j)(x_i-\bar{x})(x_j-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

公式(3)

其中，N 在时间 t 所有地区的论文产出数量，$W(i,j)$ 为空间权重矩阵；$S=\sum_{i=1}^{n}\sum_{j=1}^{n}W(i,j)$，空间权重矩阵中所有元素的和；$x_i$ 和 x_j 分别表示 i 和 j 两个地区的论文产出数量。Moran 指数在 －1 到 1 之间变化。1 表示正空间自相关，－1 表示负空间自相关，0 表示没有空间自相关。与用于描述全局空间自相关的 Moran 指数不同，局部空间自相关需要用一系列指标进行描述，这一系列指标被归纳为空间相关性局部指标(LISA)。

在进行探索性空间分析过程中，最为重要的影响因素为空间权重的确定，也会对结果产生关键性影响。笔者所使用的 GeoDa 分析工具支持基于邻接关系和基于距离的多种空间权重矩阵 $W(i,j)$ 的定义方式。由于中国行政区划并非按照经纬度划分，尤其是部分行政区面积相差甚远，基于距离的空间权重定义误差较大，因此本研究选择基于邻接关系的空间权重定义方式。

2.3.3 科研产出影响因素分析 结合空间相关性分析结果，综合讨论地区对于科研产出的影响因素，并进行相关性分析。从空间知识溢出视角出发，地区科研产出的影响因素主要有地区高校数量、国民生产总值(GDP)、地区人数以及研究与试验(R&D)投入。在相关性指标选取上，主要计算其皮尔逊相关系数(Pearson Correlation Coefficient)。其计算公式如下：

$$r_t = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{\sigma_x}\right)\left(\frac{y_i-\bar{y}}{\sigma_y}\right)$$

公式(4)

其中，$\bar{x}$、σ_x 分别为影响因素的均值与标准差。

3 研究结果分析

3.1 科研产出的地理分布情况分析

1997 年至 2016 年 20 年间以省份为观测点的科研产出(论文)分布情况如图 2 所示：

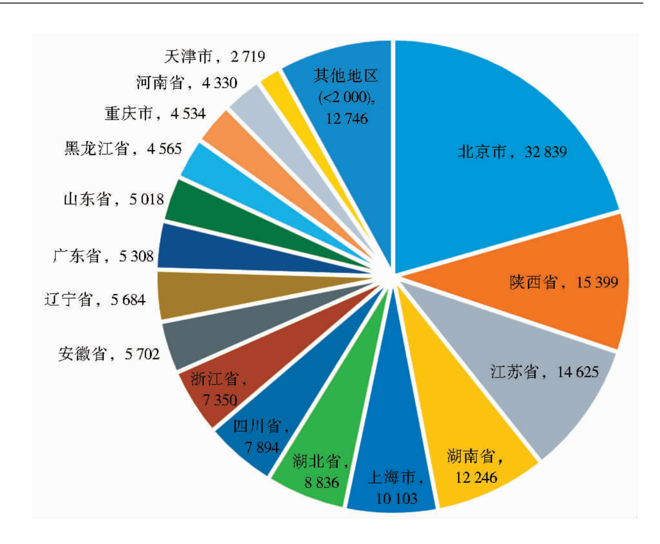


图 2 以省份为单位科研产出分布情况

从整体来看，长三角地区、北京地区以及华中地区形成了该领域期刊论文的高产地区，而西北地区则处于劣势。从分布情况来看，存在明显的不均衡性，科研产出主要集中在华北地区(北京)、长三角地区和中部地区。产出(含合著)最高的省份(含自治区、直辖市)为北京市(20.54%)、陕西省(9.63%)和江苏省(9.15%)。最低的省份(含自治区、直辖市)为西藏自治区、青海省和海南省，占比都在 0.05% 以下。

以省级行政单位为观测点大致描述了整体的分布情况，让我们对期刊论文产出在地理上的分布有了初步了解。为进一步探究内部的集中情况，笔者对观测粒度进行细化，以经纬度构建热力坐标图，观察科研产出的分布情况。1997 年至 2016 年 20 年间的科研产出(论文)分布情况热力图如图 3 所示：

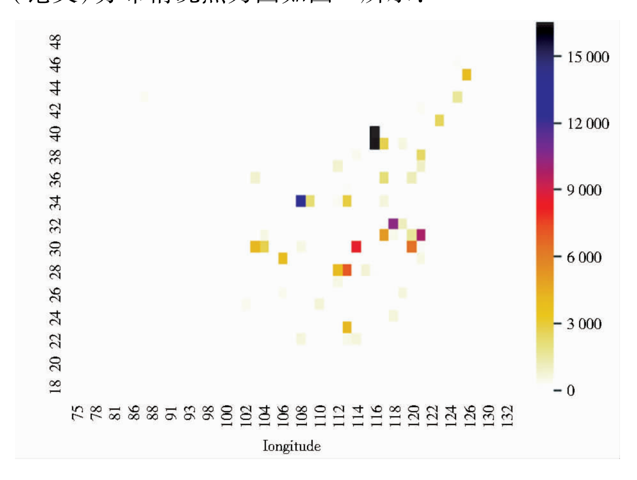


图 3 地区科研产出坐标热力图

与省份观测点进行对比，区域分布的集中情况更加直观。以城市为单位对热力图进行聚合，其中科研产出最高的城市依然为北京市，而位列前 10 位的城市均为直辖市和省会城市，其中北京市(20.54%)、西安市(7.80%)、上海市(6.32%)、成都市(4.19%)、南京市(2.89%)、重庆市(2.84%)、武汉市(2.03%)、天津市(1.70%)、广州市(1.68%)以及哈尔滨市(1.67%)。从第 12 位开始才陆续出现了大连市

(1.49%)和无锡市(0.59%)等非省会城市。并且除大连市外,其余非省会城市的产出占比均未超过 1%。同时,笔者也从数量层面进行集中程度进行测度,以年度为单位分别计算基尼指数和中心性指数,结果如图 4 所示:

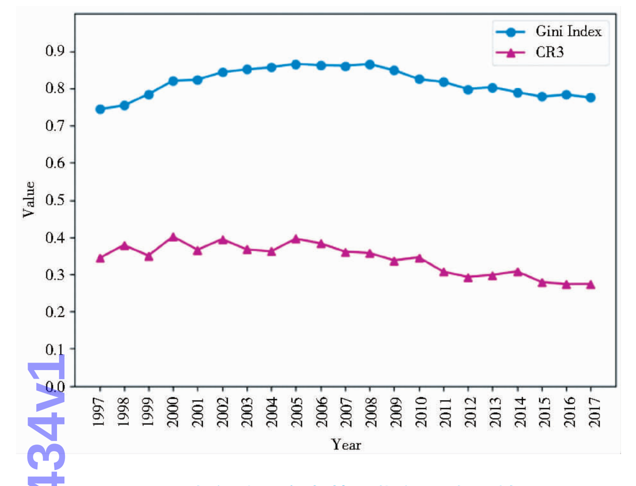


图 4 城市科研产出基尼指数和中心性指数(1997 – 2017)

从图 4 中可以发现,基尼指数始终处于 0.7 以上。最低值为 1997 年的 0.744,最高值为 2005 年的 0.866。有学者对基尼系数的合理值进行了分析,认为基尼系数在 0.33 左右是理论最佳值[20],大于 0.5 属于不均衡状态。中心性指数(CR3)的变化情况和基尼指数基本类似,前 5 年的波动较大,后续呈下降趋势,并相对平稳。基于整个时间跨度,基尼指数和中心性指数说明了期刊论文产出在空间上的不均衡性。

3.2 科研产出的空间自相关性分析

科研产出的空间分布规律以及空间聚集程度可以通过衡量科研产出的空间自相关性来研究。通过测算全局 Moran 指数,可以得到整体的空间分布规律是属于离散、随机,还是在一定程度上出现聚集,即衡量全局自相关程度。而为了更加深入的分析描述科研产出的具体分布情况,需要进行局部空间自相关的测度。与用于描述全局空间自相关的 Moran 指数不同,局部空间自相关需要用一系列指标进行描述,这一系列指标被归纳为空间相关性局部指标(LISA)[21],通过在地理空间的映射,可以直观的表现为 Moran 散点图和聚类地图,以及衡量 LISA 显著性水平的显著性地图。在对科研产出进行空间自相关分析过程中,笔者选取更细粒度的市级行政单位作为观测点,测算其空间自相关性,并对结果进行整合,最终得到以市级行政单位的空间自相关性分析结果,如图 5 所示:

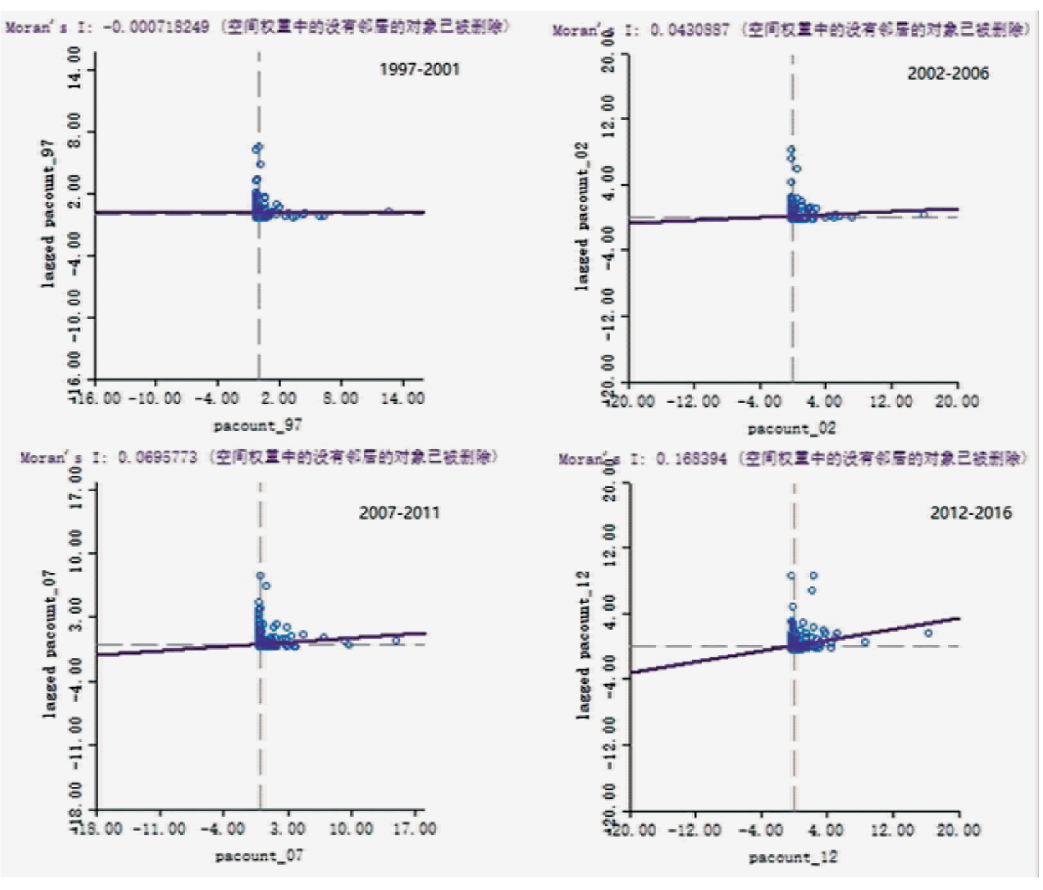


图 5 论文科研产出全局 Moran 指数散点示意

3.2.1 全局空间自相关分析结果 根据图5所示的结果，全局Moran指数大于0，从1997年到2016年20年间，期刊论文科研产出存在显著的空间自相关关系，同时，根据时间先后呈整体上升趋势。具体来看，1997-2001年5年间，全局Moran指数约等于-0.000 72，非常接近于0，因此在这5年中全局相关性并不明显。这是由于在20世纪末期，受限于交通、通信等多种因素，机构间的科研合作并未广泛开展，局限于机构内部的合作，或是独立产出，因此整体分布较为离散，无明显的空间聚集性。在未来的3个时间段中，全局Moran指数逐步升高，在2012-2016年间到达了较高值约等于0.168，有明显的自相关性，说明随着研究合作的发展，机构间的联系更加紧密，在空间上有明显的聚集效应。另外，从散点图中可以发现，有部分数据点对回归线梯度影响很大且较为分散，也从另一个角度说明了科研产出的聚集性。

3.2.2 局部空间自相关分析结果 局部Moran指数的象限分布表示了四种空间自相关关系，针对观测值，如论文发表数量，第一象限散点表示高值区域被高值区域环绕(简称高-高值聚集)，第三象限表示低值区域被低值区域环绕(简称低-低值聚集)，这两种聚集方式与全局Moran指数呈正相关关系。同理，第二象限和第四象限分别表示高-低值聚集和低-高值聚集，与全局Moran指数呈负相关关系。在进行局部空间自相关分析的过程中，笔者以市级行政区进行测算并观测局部Moran指数的象限分布，其中，有66个区域存在明显的高-高值聚集，25个区域存在明显的低-低值聚集，69个区域存在明显的低-高值聚集，还有7个区域存在明显的高-低值聚集。以市级行政区划对这些区域进行划分，局部Moran指数的空间聚集情况如表1所示：

表1 论文科研产出空间集聚特征

空间自相关关系	地区
H-H(高-高值聚集)	北京市、天津市、上海市、重庆市、武汉市、杭州市、长沙市、西安市、合肥市、成都市、济南市、沈阳市、广州市、哈尔滨市、长春市、南京市、湘潭市、大连市、青岛市、徐州市、烟台市、澳门特别行政区
L-H(低-高值聚集)	承德市、咸阳市、泰安市、苏州市、无锡市、郑州市、鄂州市、太原市、宁波市、淮南市、镇江市、张家口市、扬州市
H-L(高-低值聚集)	广元市、汕头市、钦州市、赣州市
L-L(低-低值聚集)	黑河市、白银市、兴安盟、黔东南苗族侗族自治州、邢台市、呼伦贝尔市、洛阳市、大同市、六盘水市、长治市、喀什地区、巴音郭楞蒙古自治州、鸡西市、黄山市、铜仁市、山南市、龙岩市、邯郸市、新乡市、塔城地区、鹤岗市

在图5的Moran指数散点中，各散点分别表示各市级行政区的科研产出机构数量。在实验过程中，笔者发现以地级行政区为划分时，空间自相关性不够明显，主要原因是城市之间的距离位置较远，导致了科研活动的联系较弱，同时空间权重定义采取了邻接的方式。局部空间自相关的集中趋势与总体情况分析中的科研产出占比情况基本一致，科研产出集中城市在空间上也呈现出其城市内部地区的空间聚集性。由于地区之间的地理位置接近(邻接)，使其科研产出互相影响，形成了对应的高值聚类，这与科研产出机构间的合作密切相关。那么这66个高-高值聚集区是否存在共性？为何会形成这样的聚集情况？笔者将从地区的科研产出影响因素角度进行分析。

3.3 科研产出的影响因素分析

结合前文中提及的空间知识溢出现象，已得知地区的科研产出和地区的经济发展水平有正相关性。因此，在进行地区科研产出影响因素分析的过程中，首要考虑的因素为地区经济发展水平，通常以地区国民生产总值衡量该地区的整体经济发展水平；其次，地区的科研产出和地区的科技资源投入应有相关关系[22]；另外，从所采集的数据来看，机构类型为高校的占比达到80%以上，因此地区的高等院校数量也是重要的影响因素之一；最后，笔者还考虑了地区的人口数量，主要包含两类人群：①信息传输、计算机服务和软件业从业人员数量；②科学研究和技术服务人员数量。数据来源于国家统计局，具有较强的参考性。

以省级行政单位为观测点，对地区的国民生产总值(GDP)、研究和试验(R&D)经费投入、地区高校数量、计算机服务和软件从业人员(简称行业从业人员)数量和科学研究和技术服务人员(简称科研人员)数量，与地区期刊论文产出数量(2012-2016年)进行相关性分析，结果见图6。5种因素均与地区期刊论文产出呈显著正相关关系。其中，影响力最为明显的因素为科研人员数量，平均相关系数维持在0.87以上，充分反映了科研人员在论文产出中的主力作用。相关行业从业人员的相关性次之，但平均相关系数也达到了0.85以上。行业从业人员虽然与期刊论文产出没有直接关联，但行业内的专利及著作权等，往往与科研联系紧密，从科学与技术转化的角度，也一定程度的说明了科研对地区行业发展的带动作用。地区国民生产总值以及地区的研究与试验经费投入也和地区科研产出有较高的相关性，这两个因素反映地区的经济发展水平以及发展定位，也印证了城市创新力发展投入和科研产出的相互作用性。

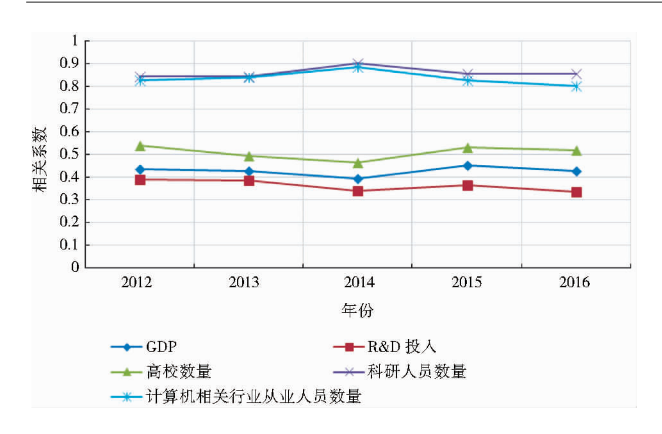


图 6 相关性因素分析(2012 – 2016)(p < 0.01)

令我们感到意外的是,高校作为科研工作的阵地,地区高校数量应与地区期刊论文产出数量存在很高的相关性,但实际的相关系数只有 0.5 左右,差强人意。主要原因是:仅从地区高校数量角度的外部因素出发,视角过于宏观与粗放。高校往往拥有多个专业,在不同领域有不同的优势与发展方向,而且从高校性质的角度出发,各高校之间的科研产出能力也有所差异,因此,仅从整体数量角度来衡量高校对期刊论文产出的影响有失公允。为了进行更加深入的分析,针对计算机软件与应用学科,笔者还结合高校该学科发展水平的实际情况,进行了更加全面探讨。基于 2017 年全国第四轮学科评估结果,笔者着重考量了计算机科学与技术以及软件工程两个一级学科的评估情况,结合地区高校分布进行了统计分析。同时,为了突出学科水平这一因素的影响,对“B –”及以上的高校数量进行累加,进行相关性分析。为方便进行比较,将其他影响因素进行了取均值处理,最终结果见图 7。统计结果证实地区高校数量与地区科研产出的相关性最高,但这个高校数量并不是绝对数量,而是针对某一领域存在优势学科的高校数量。

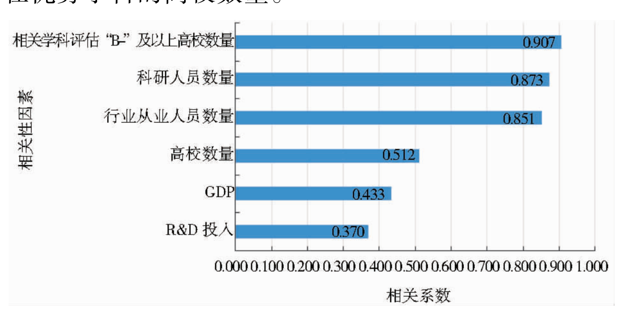


图 7 相关性因素对比(p < 0.01)

结合地区科研产出的影响因素分析结果,针对前文中得到的 66 个高 – 高值聚集地区所对应的 23 个城市,能够得出一些结论。首先,这 23 个城市涵盖了所有的直辖市。直辖市作为一种特殊的城市,属于省级行政单位,从地区规模、发展水平、人口数量等较普通城市存在明显优势。不仅拥有较高的经济发展水平,合理的科技资源配置,高等院校数量也处于领先地位,虽然在人口方面较其他省份相比不占优势,但得益于高等院校,乃至“双一流”、国内外知名高校的数量较多,其高学历、高层次人口的占比较大,这部分人口是科研产出的主要贡献力量。除直辖市外,也包含了 13 个省会城市。省会城市虽然在规模上不及直辖市,但是其汇集了本省的资源,同时大部分的省属高校基本坐落于省会城市,也形成了一定的资源聚集。除此之外,还有 5 个地级市以及澳门特别行政区存在内部的科研产出高值聚集。这些地区虽然存在科研产出,但是绝对数量并不具备优势,其形成高值聚集的原因值得探讨。观察后发现,其中大连市、青岛市和烟台市以及澳门特别行政区,这几个地区所存在的共性为均是沿海地区,由于空间权重矩阵的定义为基于邻接关系,而沿海的地区邻接关系与内陆地区有所差异。除此之外的湘潭市与徐州市,发现其城市面积较大而邻接区域较少,也导致了其形成了高值聚集。这也说明了在进行空间相关性分析的过程中,不同的空间权重矩阵定义方式将得到不同的结果。

4 总结与讨论

本文主要创新点在于:不局限于传统基于数量统计的文献计量方法,从期刊论文来源机构的地理位置信息出发,利用空间统计分析方法,揭示计算机软件与应用学科的期刊论文产出空间分布规律,并针分布情况结合多种因素进行相关性分析。

4.1 结果讨论

(1)从 1997 年到 2016 年间,计算机应用领域核心期刊地区贡献差异巨大。北京市作为我国的首都,同时也是文化中心,众多的高校数量与科研机构使其在以期刊论文为代表的科研活动中占据着无法撼动的历史地位,超过样本总量的 20%。陕西省西安市出乎意料的成为该领域期刊论文产出的高产地区,但结合地区高校数量、地区人口以及国民生产总值等影响因素,西安市并非位列前列。为此针对西安地区进行了机构层面的分析,发现西北工业大学、西安交通大学和西安电子科技大学为主要科研产出机构,在该领域贡献突出。此外,笔者还发现军事院校在该领域的研究中占有很大比重,约占样本总量的 10%,计算机应用领域,特别是图形图像相关领域,是国防科技领域的研究重

点，也充分反映了该领域对于国防科技的重要程度。同时也可以发现，不同地区机构的优势领域也有差异，不能一概而论。

（2）在研究样本的时间段内，期刊论文产出在地区之间存在显著的空间聚类效应和自相关性，且呈上升趋势，并且在直辖市以及部分省会城市存在明显的高值聚集。形成集中的主要原因有两个方面：①直辖市与省会城市的区域地位决定了其在经济发展等多方面的领先。②随着科研合作的不断发展与深入，跨机构、跨地区的论文合著成为一种有效的提升科研成果质量的合作方式，而在进行跨地区合作的过程中，地理位置接近的研究机构依然会被优先选择。这样的选择方式并非偶然，R. Boschma 等学者在 2005 年将科研合作的机构之间的相似性划分为 5 个维度[23]，同时也论证了在这 5 个维度上接近的研究机构存在更为广泛的合作行为，其中最为重要的便是物理接近，即地理位置上的接近。虽然随着通信技术、交通水平等的不断发展，跨地区合作成本逐渐降低，笔者也发现了在研究样本时间段后期集中程度存在下降趋势，但地理位置的优势仍然在跨机构合作中起主导作用。这样的近距离合作也对科研产出在空间上的聚集产生了影响。此外，不同空间权重的定义也会对自相关性分析结果产生差异。

（3）地区的科研产出与受多方面因素影响。笔者论证了地区国民生产总值（GDP）、研究和试验（R&D）经费投入、地区人口以及地区高校数量对于地区科研产出的相关性，结果符合预期。同时，结合学科评估结果，衡量地区内高校在计算机软件与应用学科范围内的学科发展水平，也更加深入地探讨了地区高校的科研能力对地区科研产出的影响。虽然地区的科研产出效率受多方面的因素影响，但是由于笔者所选取的研究样本为期刊论文，因此影响力最大的相关因素是地区高校数量。从客观角度分析，高等学校作为科研阵地，集中了大量的硕博士研究生以及科研工作者，形成了良好的科研工作氛围。而从主观角度出发，高校往往也制定相应的奖励机制或考核机制，促进科研工作者充分利用科研资源，提升科研产出效率。很多学者也进行了高校科研激励政策的相关研究，也是高等教育管理非常重要的研究问题[24]。在对地区科研产出研究中，研究与试验经费投入是一个非常重要的影响因素，但在本研究中其相关性不及地区高校数量。其原因主要是研究与试验经费投入主要影响所在地区的专利产出、著作权产出等，而这些类型的科研产出来源更为广泛，不仅局限于高等院校和研究机构。因为以城市竞争力与创新力提升发展为导向，高新技术企业也逐渐将科研产出与知识产权作为提升其综合竞争力的重要内容。在以专利为主的地区科研产出的相关研究中，地区的研究与试验经费投入、从业人员数量等都作为重要的研究维度。

4.2　研究不足与展望

本研究主要不足在于地理位置的时效性和地址解析过程中的误差控制。其中时效性问题的主要原因是研究时间跨度较长，机构可能存在名称变更、机构重组等不确定因素，降低了研究样本的质量。地址解析过程中的主要问题在于开放地图接口的解析能力不足，所得到的结果不够精确，差强人意。另外，本文中所探讨的论文产出并未涉及合作关系中的权重影响与作者优先级，在后续研究中可以综合考虑。同时从研究深度的角度出发，可以提升精度和增加维度，使结果更加丰富，得到更多有价值的信息。

参考文献：

［1］BORNMANN L, WALTMAN L. The detection of “hot regions” in the geography of science-a visualization approach by using density maps［J］. Journal of informetrics, 2011, 5(4):547－553.

［2］MATTHIESSEN C W, SCHWARZ A W. Scientific centres in Europe: an analysis of research strength and patterns of specialisation based on bibliometric indicators［J］. Urban studies, 1999, 36(3): 453－477.

［3］ZHOU P, THIJS B, GLÄNZEL W. Regional analysis on Chinese scientific output［J］. Scientometrics, 2009, 81(3):839－857.

［4］MABE M. The growth and number of journals［J］. Serials, 2003, 16(2):191－197.

［5］胡昌平，胡吉明. 领域研究合作网络的地理可视化——以《中国图书馆学报》为案例数据［J］. 情报科学，2018,36(2):3－8.

［6］TOBLER W R. A computer movie simulating urban growth in the Detroit region［J］. Economic geography, 1970, 46(S1): 234－240.

［7］DORMANN C F, MCPHERSON J M, ARAUJO M B, et al. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review［J］. Ecography, 2007, 30(5):609－628.

［8］BOTTAZZI L, PERI G. Innovation and spillovers in regions: Evidence from European patent data［J］. European economic review, 2003, 47(4):687－710.

［9］FRENKEN K, HARDEMAN S, HOEKMAN J. Spatial scientometrics: towards a cumulative research program［J］. Journal of Informetrics, 2009, 3(3):222－232.

［10］FELDMAN M P, AUDRETSCH D B. Innovation in cities: Science-based diversity, specialization and localized competition［J］.

European economic review, 1999, 43(2):409 –429.
[11] 李惠芬. 城市创新力指标体系的建构及实证研究[J]. 南京社会科学, 2010(7):15 –20.
[12] 张协奎, 邬思怡. 基于"要素 –结构 –功能 –环境"的城市创新力评价研究——以 17 个国家创新型试点城市为例[J]. 科技进步与对策, 2015, 32(2):138 –144.
[13] 曹勇, 曹轩祯, 罗楚珺, 等. 我国四大直辖城市创新能力及其影响因素的比较研究[J]. 中国软科学, 2013(6):162 –170.
[14] 史欣向,冯莉,梁彤缨. 中国现有的研发资源与科研产出的关系——基于第二次全国 R&D 资源清查数据的实证研究[J]. 科研管理,2012,33(10):1 –8.
[15] 杨军,伏琳,林艺文,等. 浅析高等学校科研投入与产出[J]. 科技管理研究, 2013, 33(16):102 –106.
[16] 吴杨,何光荣,何晋秋. 高校科研投入与产出的相关性分析: 1991 –2008[J]. 清华大学教育研究, 2011, 32(4):104 –112.
[17] 中国科学技术大学. 中国科学技术大学官方网站[EB/OL]. [2018 –11 –01]. https://www.ustc.edu.cn.
[18] AUDRETSCH D B, FELDMAN M P. R&D spillovers and the geography of innovation and production[J]. The American economic review, 1996, 86(3): 630 –640.
[19] WANG Y, HU R, LIU M. The geotemporal demographics of academic journals from 1950 to 2013 according to Ulrich's database [J]. Journal of Informetrics, 2017, 11(3):655 –671.
[20] 胡祖光. 基尼系数理论最佳值及其简易计算公式研究[J]. 经济研究, 2004, 9(1):60 –69.
[21] ANSELIN L. Local indicators of spatial association-LISA[J]. Geographical analysis, 1995, 27(2): 93 –115.
[22] 蔡翔,崔晓兰,熊静,等. 我国地区 R&D 效率及其影响因素探究——基于"科研产出 –成果转化"视角[J]. 软科学,2013,27(3):80 –84.
[23] BOSCHMA R. Proximity and innovation: a critical assessment[J]. Regional studies, 2005, 39(1): 61 –74.
[24] 李佳哲,胡咏梅. 国内高校科研效率和生产率研究述评及研究展望[J]. 现代教育管理, 2018(01):54 –61.

作者贡献说明:
马超:设计研究方案,程序编写,撰写论文;
李纲:提出研究思路与研究方法;
毛进:设计论文结构,修改论文;
谷岩松:数据预处理。

Research on the Spatial Distribution of Scientific Research Output: Taking the Computer Software and Application Field as an Example

Ma Chao Li Gang Mao Jin Gu Yansong
Center for Studies of Information Resources, Wuhan University, Wuhan 430072

Abstract: [**Purpose/significance**] The extensive development of scientific research cooperation will inevitably lead to the imbalance of regional scientific research output. In order to explore its distribution, the paper analyzed the causes of differences, and studied the spatial distribution of regional scientific research output. [**Method/process**] Based on the bibliographic data of the computer field from 1997 –2016, the geographical location information of the organization is extracted, and the observation point was used to visualize the distribution, and the Gini index and the centrality index are measured to study the distribution concentration. By measuring the global Moran index and local autocorrelation index, the spatial autocorrelation law was studied. By measuring the Pearson correlation coefficient, the correlation between the number of colleges and universities, the gross national product, the research and experimental input, and the population was studied. [**Result/conclusion**] The output of journals in the computer field showed an uneven geographical distribution, a clear concentration, and spatial autocorrelation. Meanwhile, it had a high correlation with the number of regional universities and regional research and experimental funding.

Keywords: knowledge spillover spatial autocorrelation urban innovation exploratory spatial data analysis